DreamLLM: Synergistic Multimodal Comprehension and Creation

Year: 2023 | Citations: 268 | Authors: Runpei Dong, Chunrui Han, Yuang Peng, Zekun Qi, Zheng Ge

---

Abstract

This paper presents DreamLLM, a learning framework that first achieves versatile Multimodal Large Language Models (MLLMs) empowered with frequently overlooked synergy between multimodal comprehension and creation. DreamLLM operates on two fundamental principles. The first focuses on the generative modeling of both language and image posteriors by direct sampling in the raw multimodal space. This approach circumvents the limitations and information loss inherent to external feature extractors like CLIP, and a more thorough multimodal understanding is obtained. Second, DreamLLM fosters the generation of raw, interleaved documents, modeling both text and image contents, along with unstructured layouts. This allows DreamLLM to learn all conditional, marginal, and joint multimodal distributions effectively. As a result, DreamLLM is the first MLLM capable of generating free-form interleaved content. Comprehensive experiments highlight DreamLLM's superior performance as a zero-shot multimodal generalist, reaping from the enhanced learning synergy. Project page: https://dreamllm.github.io.